

THE FORCED LABOR STRUCTURAL RISK INDEX

A COUNTRY-LEVEL MEASURE OF THE CONDITIONS UNDER WHICH FORCED LABOUR BECOMES MORE LIKELY

Ethical Tech CoLab
NYU Center for Global Affairs
December 2025

Prepared as masters research at the NYU Center for Global Affairs.

Forced labour is one of the few large-scale human rights violations that almost nobody can count, and the countries where the risk is gravest are frequently the countries where the evidence is thinnest, because the same institutional weakness that permits exploitation also prevents it from being recorded. FLSRI does not try to estimate how many people are in forced labour. It measures something different and, for prevention purposes, more tractable: how strongly the structural conditions that make forced labour more likely are present, and how those conditions combine.

KEY FIGURES

184

countries of roughly 195 receive a composite score; the rest are reported as not scored rather than given an invented value

43

standardised indicators across eleven domains, each drawn from a named cross-national dataset

10,000

simulated re-scorings behind the uncertainty band published for every scored country

45

ranks wide is the median uncertainty band, which is why the index asks to be read in tiers rather than as a league table

01

EXECUTIVE SUMMARY

FLSRI is a research prototype: a country-level index that scores the structural conditions under which forced labour becomes more likely. It scores 184 of roughly 195 countries on a scale running from 0 to 1, where 0 is the lowest modelled structural risk and 1 the highest. The remaining countries are reported as not scored rather than being given an invented value.

The index is deliberately not a prevalence estimate. It does not say how many people are exploited. A high score means the enabling conditions hold together strongly; a low score does not certify a country as free of forced labour. This distinction is stated repeatedly in the project's own documentation and it governs every legitimate use of the results.

The framework rests on a simple structural claim borrowed from criminology: exploitation becomes likely when a population exposed to coercive recruitment coincides with an environment in which exploiting that population goes unchecked. The index therefore has two scored halves. Recruitment, written as R, asks who is structurally exposed, through poverty, debt, blocked mobility, exclusion, weak legal identity, gendered labour structures, childhood exposure, and shocks such as conflict and disaster. Exploitation, written as E, asks whether exploiting them can run

without consequence, through blocked exit, demand in high-risk sectors, and the state's own production of unfreedom.

The two halves are combined by a geometric mean, that is, the square root of R multiplied by E . This is a deliberate methodological choice. It means a country can only score high when both halves are high. A very exposed population in a country with strong labour enforcement, or a permissive enforcement environment in a country with little structural vulnerability, is pulled down rather than averaged out.

A third phase, Monetization, covering the financial conditions under which the proceeds of forced labour can be moved, hidden, and kept, is computed and published but deliberately excluded from the headline score. It answers a different question, namely where intervention against the money would bite, and it scores high for wealthy financially opaque economies in a way that would distort a structural risk reading.

Eleven domains and forty-three standardised indicators feed the published score, each drawn from an established cross-national dataset such as the World Bank World Development Indicators, ILOSTAT, the V-Dem democracy dataset, the UNHCR population statistics, the UCDP conflict event dataset, and the EM-DAT disaster database.

The project is unusually disciplined about two failure modes that afflict composite indices of this kind. The first is circularity, that is, predicting forced labour using a measure of forced labour. The second is the risk that the index becomes a governance ranking wearing a new label. Both are addressed explicitly, tested, and reported rather than concealed.

Every scored country carries a published uncertainty band derived from 10,000 simulated re-scorings, and the documentation insists that the results be read in broad tiers rather than as an exact league table. The middle of the table is openly described as unstable. The index is delivered as an interactive public website together with the complete pipeline that produces it, so that any figure shown can be traced back to the data and the code that generated it.

BACKGROUND AND RATIONALE

The problem. Forced labour is defined in international law and prohibited almost universally, yet it is measured very unevenly. Global estimates exist, but they are produced from surveys and administrative sources that are not available for every country, are refreshed slowly, and were never designed to support a fine-grained country-by-country comparison. Detection statistics, such as counts of identified trafficking victims, are worse still for this purpose: they largely record how much capacity a state has to detect and register cases, so a well-resourced state can appear worse than a state where nothing is investigated at all.

The gap. Prevention and triage need to know where conditions are dangerous before cases surface. Existing vulnerability measures go some way toward this, but several of them are built partly from the same governance material that dominates most cross-national comparison, which makes it hard to tell whether they are measuring forced-labour risk or simply restating institutional weakness.

The response. FLSRI proposes a purpose-built structural measure. Its guiding commitments, visible throughout the code and documentation, are that every figure should be traceable to a named public dataset, that missing information should be declared rather than filled in, that the reasoning behind each methodological choice should be written down, and that the limits of the result should be published as prominently as the result itself.

The project describes itself as publishing a framework, a pipeline, and code, with the country scores presented as estimates of structural conditions carrying explicit uncertainty. It does not present the country rankings as findings about the world.

03

OBJECTIVES

The index is designed to do the following.

- Measure the structural conditions associated with forced labour across almost all countries, on a single comparable scale, without estimating prevalence.

- Make the underlying theory explicit, by separating exposure to recruitment from the conditions for unchecked exploitation and requiring both to be present before a country scores high.
- Ground every indicator in a real, citable, cross-national dataset, with the source, vintage, coverage, licence, and required citation recorded for each one.
- Refuse to fabricate. Missing data is never silently converted to zero, and countries whose evidence base is too thin are left unscored.
- Guard against circularity by excluding inputs that are themselves measures or detections of forced labour and trafficking, wherever using them would mean predicting forced labour with a measure of forced labour.
- Report, rather than engineer away, the index's association with weak governance, while demonstrating that the index is not reducible to a governance ranking.
- Publish uncertainty alongside every score, and direct readers toward tiers rather than exact ranks.
- Support prevention and triage by identifying where structural risk concentrates, including below the national level, and where in the structure a government, a buyer, or a non-governmental organisation might act.

04

THE LEGAL AND NORMATIVE FRAME

Forced labour has a settled definition in international law. The International Labour Organization's Forced Labour Convention, 1930 (No. 29), Article 2(1), defines forced or compulsory labour as all work or service which is exacted from any person under the menace of any penalty and for which the said person has not offered himself voluntarily. The Abolition of Forced Labour Convention, 1957 (No. 105), prohibits the use of forced labour for specified purposes including political coercion, labour discipline, and discrimination. The 2014 Protocol to Convention No. 29 obliges ratifying states to take effective measures on prevention, on the protection of victims, and on access to remedies including compensation. It entered into force in November 2016.

To help identify forced labour in practice, the ILO publishes a set of eleven indicators of forced labour, first issued in 2012 and reissued in a revised edition in 2025. The eleven are these.

- Abuse of vulnerability.
- Deception.
- Restriction of movement.
- Isolation.
- Physical and sexual violence.
- Intimidation and threats.
- Retention of identity documents.
- Withholding of wages.
- Debt bondage.
- Abusive working and living conditions.
- Excessive overtime.

The ILO describes their purpose as assisting law enforcement officials, labour inspectors, trade union officers, and non-governmental organisation staff to identify people who may be trapped in forced labour and may need urgent help.

The ILO is careful about how much weight any one indicator carries. Its guidance states that the presence of a single indicator may in some cases imply the existence of forced labour, while in other cases several indicators must be read together before a case can be said to exist. The eleven are a screening aid for practitioners, not a legal test, and they are frequently over-read as a checklist. That caution matters for an index built on the same vocabulary.

These are operational indicators for recognising a case in front of you. FLSRI works at a different level: it measures the standing country conditions that make those case-level indicators more likely to arise, and its documentation states that its domain set is broadly aligned with these indicators and with the vulnerability dimensions used in the Walk Free Global Slavery Index.

The obligation is not only a state obligation. The United Nations Guiding Principles on Business and Human Rights, endorsed by the UN Human Rights Council in 2011, rest on three pillars: the state duty to protect human rights, the corporate responsibility to respect them, and access to remedy for those harmed. The Guiding Principles on human rights due diligence require enterprises to identify, prevent, mitigate, and account for how they address their human rights impacts, including impacts in their supply chains

and business relationships. A structural risk measure is directly relevant to that first step: an enterprise or a public procurement body cannot prioritise due diligence without some defensible view of where risk is concentrated.

Target 8.7 of the Sustainable Development Goals commits states to take immediate and effective measures to eradicate forced labour, end modern slavery and human trafficking, and secure the prohibition and elimination of the worst forms of child labour. Two of the index's own indicators are drawn directly from the SDG monitoring framework: informal employment as a share of total employment, which is SDG indicator 8.3.1, and the completeness of birth registration, which is SDG indicator 16.9.1.

The index does not claim to establish a violation of any of these instruments, and it should not be cited as evidence that a particular state is in breach of its obligations. It is a prioritisation instrument for prevention, not a compliance finding.

05

HOW THE INDEX IS BUILT

The index has a fixed shape with four levels: indicator, domain, phase, and composite. Each level sits on the same 0 to 1 scale.

The indicator. An indicator is a single measured quantity taken from a real dataset, for example the share of the population living below the 6.85 dollar a day poverty line, or the number of labour inspectors per ten thousand employed people. Each one arrives in its own units, so the pipeline rescales it onto the common 0 to 1 risk scale.

The domain. A domain is a group of indicators measuring one mechanism, for example economic precarity or constrained mobility. The domain score is the plain average of the indicators that are present for that country. Indicators that are missing are dropped from the average and are never entered as zero.

The phase. A phase is one side of the structural claim. Recruitment contains eight domains; Exploitation contains three. A phase score is the plain average of the domains scored inside it. The eleven are set out below and described one by one in the next section.

PHASE	DOMAIN	MECHANISM IT IS MEANT TO CAPTURE
Recruitment	Economic precarity	Needs that the available work cannot meet
Recruitment	Debt and financialised dependency	Obligation that turns a voluntary arrangement into an inescapable one
Recruitment	Constrained mobility	Inability to move legally or affordably
Recruitment	Ascriptive exclusion	Exclusion by group membership rather than by conduct
Recruitment	Legal non-recognition	Inability to prove who you are and so to claim protection
Recruitment	Gendered labour	Concentration of women in exploitation-exposed work
Recruitment	Age and childhood structuring	Direct exploitability of children and household labour exhaustion
Recruitment	Structural disruption	Conflict, displacement, and disaster shocks
Exploitation	Foreclosed exit	The cost of walking away from a bad situation
Exploitation	Economic structure and demand	How much of the economy sits in high-risk work
Exploitation	State production of unfreedom	Unfreedom generated by legal architecture and impunity

The eleven scored domains and the phase each belongs to. The Monetization lens is computed separately and excluded from the headline score.

The composite. The published score is the geometric mean of the two phases, that is, the square root of Recruitment multiplied by Exploitation. If either phase cannot be scored for a country, the composite is not scored. There is no substitution of zero for a missing phase.

Indicators are rescaled by what the project calls absolute anchoring. For each indicator, a floor value and a ceiling value are fixed in advance. The floor is the raw value that maps to 0 and the ceiling is the raw value that maps to 1; anything beyond either end is clamped.

INDICATOR	FLOOR, MAPPING TO 0	CEILING, MAPPING TO 1
Share of employment in agriculture	0 per cent	80 per cent
Gender gap in labour force participation	0 percentage points	50 percentage points
Conflict deaths	0 per 100,000 people	100 per 100,000 people

Three examples of the anchor ranges the pipeline uses.

Anchoring to fixed reference points rather than to the best and worst countries observed is a deliberate choice. It means a country's score does not move simply because a different set of countries happened to have data this year, and it makes scores comparable across refreshes of the underlying data. Where a quantity depends on the size of a country, it is first converted to a rate, a share, or a gap before scaling, so that a large country is not scored as riskier merely for being large.

Each indicator carries an explicit direction. Most point the same way as risk: more poverty means more risk. Some are protective and are inverted, so that a high raw value becomes a low risk score. Trade union density, collective bargaining coverage, labour inspector density, freedom of movement, and visa-free passport access are all protective indicators entered in inverted form. By the time any indicator reaches the scoring stage, every value points in the same direction, with higher meaning more risk.

The documentation is candid that several anchors were set from the observed distribution, for instance a ceiling near the 95th percentile, which reintroduces a degree of dependence on the particular sample of countries into a scale that is presented as absolute. Locking these anchors and testing how much the results move if they shift is listed as outstanding work.

Every level uses equal weights. Indicators inside a domain are averaged equally, domains inside a phase are averaged equally, and the two phases enter the geometric mean equally. The documentation is unusually direct that this is a value judgment rather than a neutral default. Equal weighting asserts that each component is equally relevant to the construct, which is a substantive claim and not an absence of one.

Equal weighting at each level produces unequal weight per indicator once the whole structure is taken into account, because the boxes are not the same size. A domain containing one indicator gives that indicator the full weight of a domain, while a domain containing five splits it. More consequentially, the Exploitation phase contains three domains against Recruitment's eight, so a single Exploitation domain carries roughly 2.7 times the marginal weight of a single Recruitment domain. The project's own interactive material states this plainly rather than leaving it to be discovered.

The project reports that shifting the balance between the two phases as far as 60 to 40 in either direction barely moves the ranking. It also states that it has not yet published a similar test for perturbing weights at the indicator and domain level, and records this as a known gap.

The governing principle on missing data is that it is never silently treated as zero. A missing indicator is dropped from its domain average rather than counted as an absence of risk. Dropping cannot go on indefinitely, so a coverage floor applies. A domain is scored only if at least half of its mapped indicators are present and never fewer than two. The same floor applies to the count of scored domains within a phase. Below the floor, the domain or phase is marked not scored. A domain that is designed around a single indicator cannot mechanically meet the requirement for two; these are handled as a named exception, scored if the one indicator is present but always carried as low confidence.

Under these rules, 184 of roughly 195 countries receive a composite score. The eleven unscored countries are Andorra, Dominica, the Federated States of Micronesia, Saint Kitts and Nevis, Liechtenstein, Monaco, the Marshall Islands, Nauru, San Marino, Tuvalu, and the Holy See. The documentation notes that this missingness is not random: these are micro-states and small island states for which the labour and governance series are simply not collected. It also warns that within the scored set, thin coverage tends to push a score down rather than up, so a very low score for a data-sparse state deserves extra caution.

THE VARIABLES, EXPLAINED IN PLAIN TERMS

This section is the heart of the report. It sets out, domain by domain, what each indicator represents, why it was chosen, and how it enters the score. All indicators are averaged equally within their domain. Phase R, Recruitment, asks who is structurally exposed, and holds eight domains.

Economic precarity. The proposition is that people who cannot meet their needs through the work available to them accept terms they would otherwise refuse. Five indicators carry it. Poverty headcount is the share of the population living on less than 6.85 dollars a day, from the World Bank World Development Indicators, anchored from 0 to 100 per cent. Informal employment share is the proportion of total employment that is informal, taken from ILOSTAT as SDG indicator 8.3.1; informal work sits outside labour inspection, contract enforcement, and social protection, which is precisely the exposure the domain is trying to capture. Agrarian employment share is the proportion of employment in agriculture, anchored from 0 to 80 per cent, because agriculture concentrates seasonal, isolated, and poorly regulated work. Income volatility is the standard deviation of real GDP growth over roughly fifteen years, anchored from 0 to 8 percentage points, on the reasoning that instability and not only low income is what pushes households into distress decisions. Income inequality is the Gini index, anchored from 20 to 65.

Debt and financialised dependency. Debt is the classic mechanism by which a voluntary arrangement becomes an inescapable one, and debt bondage is one of the ILO's eleven case-level indicators. All three measures come from the World Bank Global Findex Database, a large survey of how adults around the world save, borrow, and make payments: the share of adults with no financial account, the share who borrowed from informal sources in the past year, and the share who borrowed any money in the past year. The domain is carried as low confidence, and the reason is stated openly. The mechanism the designers actually wanted, namely recruitment-fee and migration debt, has no cross-national source, so the domain rests on financial exclusion and borrowing patterns instead. Borrowing prevalence in particular measures how widely credit is used, not whether households are in distress. The borrowing questions cover only about 59 per cent of countries.

Constrained mobility. The idea is that people who cannot move legally or affordably are easier to recruit on bad terms and harder to extract from a bad situation. Freedom of movement comes from V-Dem, an expert-coded dataset

on the qualities of political regimes, entered inverted. Passport access is the number of destinations reachable without a visa, from the Henley Passport Index, also inverted, on the reasoning that a weak passport means legal migration routes are scarce and irregular ones become the alternative. Refugees originating from the country, per 100,000 people, come from the UNHCR population statistics. A tied-status coding is derived from the DEMSCORE legal-status data, covering 29 countries, alongside a hand-coded kafala tied-status signal covering eight sponsorship-system states.

The kafala signal. Under sponsorship systems used across parts of the Gulf and the wider region, a migrant worker's legal status is tied to a single employer, so leaving that employer can mean losing the right to remain. The index codes three features for the eight adopting states it covers: whether status is tied to one employer, whether the employer's consent is required to leave, and whether leaving is treated as an offence. This is a narrow, hand-built input covering a handful of countries, not a global series, and the project is explicit that the sponsorship mechanism remains unsourced at country scale. The domain is carried as low confidence.

Ascriptive exclusion. Exclusion by group membership, meaning by who a person is rather than what they have done, is captured by a single indicator: the depth-weighted share of the population belonging to politically excluded ethnic groups, from the Ethnic Power Relations dataset maintained at ETH Zurich, which codes ethnic groups' access to executive power. Two limitations are recorded. The indicator captures political exclusion, not exclusion within the labour market, which is the channel most relevant to forced labour. And roughly 41 countries lie outside the dataset's universe and are left missing rather than scored as zero. The domain is single-indicator and therefore always carried as low confidence.

Legal non-recognition. A person who cannot prove who they are cannot easily enforce a contract, claim a wage, access protection, or cross a border lawfully. Birth-registration incompleteness is derived from the completeness of birth registration reported through the World Bank from UNICEF and UN Statistics Division civil registration data, corresponding to SDG indicator 16.9.1, entered inverted. Statelessness prevalence, in persons per 100,000, comes from the UNHCR population statistics. Where statelessness figures are suppressed or absent, the domain rests on birth registration alone; 26 countries are in that position and are flagged as low confidence.

Gendered labour. Forced labour is strongly gender-patterned, particularly in domestic work, care work, and sectors where women are concentrated. Four indicators carry the domain: the gap between male and female labour force

participation, in percentage points, anchored from 0 to 50; the UNDP Gender Inequality Index, covering reproductive health, empowerment, and labour market participation, running from 0 for equality to 1 for maximum inequality; legal constraints on women's mobility, taken from the mobility component of the World Bank Women, Business and the Law project, entered inverted; and sex-by-sector channelling from ILOSTAT, the extent to which one sex is concentrated in exploitation-exposed sectors of employment. The documentation is careful to record that a global prevalence estimate is used only to justify why this mechanism matters, and is never scored. The value entered is always a structural employment share, never a victim count.

Age and childhood structuring. Children are both directly exploitable and an indicator that households are already using every available source of labour. Four indicators, all from the World Bank's mirrors of ILO and UNICEF SDG series: child labour prevalence, being the share of children aged 7 to 14 in employment; the out-of-school rate for lower-secondary-age adolescents; the share of the population aged 0 to 14; and the child marriage rate, the share of women aged 20 to 24 first married before 18. This domain carries the project's most awkward data problem. Child labour prevalence, the indicator that most directly measures the mechanism, covers only about 47 per cent of countries, from survey vintages spanning 2005 to 2016. That is below the coverage floor. Consistent with the no-imputation rule, the gap is disclosed rather than filled. The documentation also records that the causal link from child marriage to forced labour is under review rather than settled.

Structural disruption. Shocks displace people, break household economies, and hand recruiters a supply of people with nothing to fall back on. Disaster-affected intensity comes from EM-DAT, maintained by CRED at UCLouvain, measured over the five years 2020 to 2024, with heat-wave entries excluded on the documented ground that heat mortality is recorded only where a country runs attribution studies, so it tracks measurement capacity rather than shock. Climate vulnerability comes from the ND-GAIN Country Index, using only the vulnerability axis; ND-GAIN's separate readiness axis is left out to avoid re-importing governance. Conflict intensity, in deaths per 100,000 over the five years 2019 to 2023, comes from the Uppsala Conflict Data Program's Georeferenced Event Dataset. Internally displaced persons per 100,000 and refugees originating from the country per 100,000 both come from UNHCR. The last of these also appears in constrained mobility, and the documentation flags this shared use as a possible double-count awaiting a correlation check.

Phase E, Exploitation, asks whether exploitation can run unchecked, and holds three domains.

Foreclosed exit. This domain is meant to capture the cost of walking away: whether a worker in a bad situation has any realistic alternative. Its three indicators are all from ILOSTAT and all protective, entered inverted: labour inspectors per ten thousand employed persons, the share of employees covered by collective agreements, and trade union density. The project is candid that this domain does not measure what it was designed to measure. The intended mechanism, the cost of exit and the degree to which a single employer dominates a local labour market, cannot be sourced across nearly 200 countries. What is available are protective factors whose absence is being read as risk, which is not the same thing. Coverage runs from about 43 to 69 per cent. The domain is carried as insufficient data and is explicitly labelled a stand-in; whether it should be scored at all, folded into a neighbouring domain, or held aside as a diagnostic, remains an open question. Because Exploitation has only three domains, this weakness matters more than it would in Recruitment.

Economic structure and demand. Some economies simply have more of the work in which forced labour occurs. Three indicators: hazardous-sector share, using employment in agriculture, anchored from 0 to 60 per cent; informal employment share, from ILOSTAT; and export concentration, from UNCTADSTAT, a measure of how concentrated a country's merchandise exports are in a small number of products, used here as a proxy for concentrated buyer power. Two caveats are recorded. The export concentration measure is a low-confidence proxy for the buyer concentration the designers actually wanted. And the criminal-market side of demand has no sourced indicator at all, because the obvious candidate, organised-crime outcome measures, was excluded as circular. Agriculture and informality also appear in Recruitment, a cross-phase overlap that the documentation names as an unresolved issue.

State production of unfreedom. The proposition is that states can generate unfreedom directly, through the legal architecture they impose and the impunity they permit. The published build scores it from clientelism, from V-Dem, capturing the exchange of goods and favours for political support; bribery risk, from the TRACE International Bribery Risk Matrix; firm bribery incidence, from the World Bank, being the share of firms reporting at least one request for a bribe; the forced-labour share of detected trafficking victims, derived from the UNODC Global Report on Trafficking in Persons, covering 142 countries; and the DEMSCORE tied-status coding together with the eight-country kafala coding. Two things must be said about this domain. First, only two of five designed drivers are sourced, which is why it is

carried as insufficient data; the tied-status, immigration-architecture, and protective-floor legal codings that the design calls for do not exist as global series. Second, the domain is where the index comes closest to being a corruption measure, which is exactly why it receives the special treatment described in the next section.

The Monetization lens, computed but excluded. A third phase covers the financial conditions under which the proceeds of exploitation can be moved, hidden, and retained. It has two domains. Transnational concealment uses the FATF mutual-evaluation effectiveness score and grey-list or black-list standing, both taken from the Basel AML Index published by the Basel Institute on Governance, together with the Tax Justice Network Financial Secrecy Index. Cash and informal retention uses the World Bank Informal Economy Database estimate of the shadow economy as a share of official GDP, and the share of adults without a financial account. This phase is computed, published, and deliberately kept out of the headline score. The reason given is that it answers an intervention question, namely where the money could be disrupted, rather than a structural risk question, and that it scores high for wealthy, financially opaque economies in a way that would distort the risk reading. The project reports having tested whether any governance-independent part of the lens deserves inclusion, and found that the remainder carried no forced-labour-specific signal and diluted the composite. The lens therefore remains display-only. The project frames this distinction as one between drivers and disruptors: Recruitment and Exploitation generate risk, while Monetization is where the cycle could be broken without first removing the underlying vulnerability.

07

GUARDING AGAINST CIRCULARITY AND THE GOVERNANCE PROBLEM

Circularity. An index that predicts forced labour using a measure of forced labour tells you nothing. The project applies this discipline in several places. The United States Trafficking in Persons report, organised-crime outcome measures, and the Basel AML composite were excluded as circular. Detection counts of trafficking victims were refused as a validation benchmark on the ground that detection reflects state capacity rather than prevalence, so matching it would reward strong states. In the published build, the V-Dem indicator measuring freedom from forced labour was dropped entirely from the scoring, on the ground that using a freedom-from-forced-labour measure to predict forced-labour risk is circular by construction.

The discipline is not applied perfectly, and a careful reader should notice the tension. Detection data is refused as a benchmark for checking the index, yet one derived detection measure, the forced-labour share of detected trafficking victims from the UNODC reporting, does enter the scoring of the state production of unfreedom domain. The share of detected cases is not the same quantity as the count of detected cases, and it is less directly a function of state capacity, so the two positions can be reconciled. But the reconciliation is not spelled out in the documentation, and it is the one place where the project's own circularity rule is applied less strictly to an input than to a benchmark.

The governance problem. Most cross-national indicators correlate with the quality of a country's institutions. An index assembled from such indicators can easily become a rule-of-law ranking with a new name, which would make it uninformative for the specific question at hand.

The published response is called de-biasing. Each indicator was tested for how much of its variation is explained by a standard rule-of-law measure. Any indicator explained at or above 55 per cent was treated as too entangled with governance and entered at half weight, applied within the state production of unfreedom domain, which is where such indicators concentrate. This is a targeted reduction rather than a removal.

The result is reported honestly rather than presented as solved. After de-biasing, rule of law still explains about 63 per cent of the variation in the composite, corresponding to a correlation of about 0.79. The project's position is that this residual is correct rather than an artefact: weak governance is a genuine structural driver of forced-labour risk, so a substantial but bounded association is what one should expect. Roughly a third of the variation is not governance.

That position was stress-tested. The de-biasing was re-run varying both its scope, applying it to the one domain or to all domains, and the reference measure used to flag entangled indicators, using the World Bank rule-of-law measure, the V-Dem rule-of-law measure, or an equal blend of the two. Across every combination, rank agreement with the published index stayed at or above 0.96 on Kendall's measure of rank correlation, the governance share stayed near 63 per cent, the top five and bottom five countries were identical, and the result for the Gulf states was unchanged. The check can be re-run by anyone with the repository. This is the strongest available answer to the objection that the index is governance with a new label, and it is a more thorough answer than most composite indices provide.

One nuance is worth noting for readers who look at the code. An older and simpler mechanism, in which a domain score was reduced in proportion to a governance dial, survives in a secondary reproduction build that the repository also ships. It is not the mechanism behind the published figures. The documentation distinguishes the two carefully, and also notes that this secondary build is more governance-dominated than the published one.

08

READING THE RESULTS

The scale. Scores run from 0 to 1. In the published build the scored range is narrower than the theoretical one, running from about 0.11 to about 0.67. The highest-scoring countries are Yemen, Chad, Afghanistan, South Sudan, and Sudan; the lowest are the Nordic countries, Iceland, and several Western European states. This ordering is a face-validity check, not a discovery.

Read tiers, not ranks. The index is banded into three tiers using cuts at 0.281 and 0.402, giving 59 countries in the higher tier, 60 in the middle, and 65 in the lower tier. The cut points were fixed at the registration stage and deliberately not re-derived once the results were known, on the principle that banding thresholds must not be tuned after seeing the output. The project notes that the current build's own thirds would fall at 0.274 and 0.397, and that nine boundary countries would change tier if the cuts were re-derived.

Uncertainty bands. Every scored country carries a rank band produced by running the whole scoring 10,000 times with small random disturbances added to its two phase scores and with the balance between the phases redrawn each time between 40 and 60 per cent. The band reported is the range within which the country's rank fell in 90 per cent of those runs.

These bands are the single most useful discipline the index publishes. The median band is 45 ranks wide, and only about 2 per cent of countries have a band narrower than 10 ranks. In plain terms, a mid-table country's exact position is close to meaningless, while the extremes are stable: the top decile retains about 79 per cent of its members across the simulations, and a country stays in its published tier in about 84 per cent of runs on average.

The low-confidence badge. Forty scored countries have two or more of the eleven domains unscored. These carry a visible lower-confidence marker on

the site, and the same rule widens their uncertainty band in the simulation. One rule drives both statements, which prevents the presentation and the arithmetic from drifting apart. The uncertainty code notes that an earlier version of this rule was applied incorrectly, silently widening the band for every country, and records the date the error was corrected.

What a score is not. A high score does not mean a specified number of people are exploited. A low score does not certify a country as free of forced labour. A difference of a few places in the middle of the table is not a finding.

09

BEYOND THE NATIONAL AVERAGE

A national score averages an entire country into one number, and that average hides the regions where risk actually concentrates. The project therefore publishes a sub-national layer covering first-level administrative regions, built from census microdata from IPUMS-International, an archive of harmonised census samples from around the world, with disaster exposure taken from the Geocoded Disasters dataset.

The published surface covers 1,735 first-level administrative units, of which 1,454 pass reliability filtering and carry a risk score. Roughly 17 per cent of the variation in the underlying precarity measure sits within countries rather than between them, which is the quantitative form of the argument that national averages conceal a great deal.

The project also publishes a corridor view, on the reasoning that conditions enabling forced labour frequently span neighbouring states and that a response confined to one country addresses only part of a cross-border problem.

These layers are best read as illustrative rather than authoritative. They rest on a restricted-access data source that cannot be redistributed with the repository, and they cover fewer countries than the national index.

10

VALIDATION

The project registered its validation criteria in advance, including the thresholds at which each test would be judged to have failed, and then re-ran them on the build that the site actually displays. Pre-registration matters here: it prevents the criteria from being adjusted after the results are known.

The reported outcomes on the displayed build are as follows. Rule of law explains 0.628 of the composite's variation, passing the pre-set requirement that it stay at or below 0.80. The two phases correlate at 0.659, which is within the pre-set window of 0.30 to 0.90 and supports the claim that they are related but not redundant. A forced-labour-specific signal, child labour prevalence drawn from an independent source, remains significantly associated with the composite once governance is held constant on both sides. The top decile retains 86.5 per cent of its members under the suite's mild-noise test, against a requirement of 70 per cent. The published uncertainty model, which disturbs the inputs more aggressively, gives a lower figure of about 79 per cent for the same quantity; both are reported.

One criterion was not met, and the project reports it rather than dropping it. Tested against an external prevalence estimate, the Walk Free Global Slavery Index country prevalence figures, with governance netted out on both sides, no significant association was demonstrated. The project's own reading is that this null result is uninformative rather than disconfirming, because the benchmark is itself heavily entangled with governance. That is a defensible reading, but it is also an admission that the index has no clean external check.

The absence of a governance-independent measure of forced labour prevalence is precisely the gap that motivates a structural index, and it is also the reason the index cannot be fully validated. Readers should hold both facts at once.

A note on what was validated. The tests examine the structure: whether the index is distinguishable from governance, whether its two halves are distinct, whether it carries forced-labour-specific information, and whether its extremes are stable. They do not validate any individual country's score.

LIMITATIONS AND CAVEATS

The project's credibility rests substantially on the candour of this list, which is reproduced here in the terms the repository itself uses.

It measures conditions, not cases. The index does not estimate prevalence and cannot be read as a count of victims.

It reads origin-side risk and under-reads destination systems. This is the most consequential limitation. The mechanisms that drive forced labour in destination economies, namely sponsorship systems that tie a worker's legal status to one employer, recruitment-fee debt, and migration brokerage, are named in the framework but are not sourced at country scale, beyond the narrow hand-coded signal covering eight states. The available indicators describe resident citizens rather than the migrant workforce most exposed. As a direct result, the United Arab Emirates ranks 146th of 184 scored countries and the other Gulf states sit near it, despite well-documented risk in their labour systems. A low score for a known destination country means that this index does not yet capture that pathway, not that the country is clear. The project mitigates this by attaching a caveat banner to the eight sponsorship-system country profiles, which makes the blind spot visible where a reader would otherwise be misled.

It is correlated with weak governance by design. About two-thirds of the variation is shared with rule of law. The relationship is disclosed and stress-tested rather than engineered away, but a reader who wants a measure independent of institutional quality will not find one here.

The Exploitation phase is thin. One of its three domains does not measure its intended mechanism at all and is carried as insufficient data. Another leans partly on corruption proxies. Because the phase has only three domains, each carries roughly 2.7 times the marginal weight of a Recruitment domain, so these weaknesses propagate further than they would elsewhere.

Some indicators do double duty. Informality and agricultural sector share enter both phases, and refugee outflow enters two Recruitment domains. The collinearity screen that would resolve this is flagged in the rules but has not been run and reported.

Equal weighting is a substantive claim. It has not been tested by perturbation at the indicator and domain level, only at the phase level.

Several anchors were derived from the observed distribution. They were not derived from theory, which weakens the claim that the scale is absolute. An anchor-shift sensitivity test is outstanding.

Coverage is uneven and the gaps are not random. The child labour indicator, which measures the mechanism most directly, covers under half the countries and draws on survey vintages up to two decades old. The unscored countries are systematically small states. Thin coverage tends to depress rather than inflate a score.

Some inputs are licence-restricted. Several sources cannot be redistributed with the repository, and several others carry re-publication flags that the documentation itself marks as unconfirmed, including the Henley Passport Index, the TRACE Bribery Risk Matrix, the UNDP Gender Inequality Index republication, and the Tax Justice Network Financial Secrecy Index. EM-DAT and IPUMS-International are recorded as redistribution-restricted and are not bundled. Anyone republishing this material should resolve those questions first.

The documentation does not describe the published scoring quite completely. The generated codebook lists the indicators drawn from the main data pipeline and states that it cannot drift from that pipeline, but the published scorer additionally enters three inputs held outside it: the UNODC detected-victim composition measure, the DEMSCORE tied-status coding for 29 countries, and the kafala tied-status coding for eight countries. These are listed on the site's own indicator data and are therefore visible to a reader who looks, but they are absent from the codebook, and the narrower coverage of two of them is not reflected in the domain confidence flags. A future revision should bring the codebook and the published scorer back into alignment.

The mid-table is not readable as a ranking. With a median uncertainty band of 45 ranks, ordinal comparisons between similarly placed countries are not supportable.

There is no clean external benchmark. The index cannot be checked against a governance-independent measure of forced labour prevalence, because none exists.

It is a research prototype. It was produced as masters research, it is published as a framework and a pipeline rather than as authoritative country figures, and it should not be used as the sole basis for any consequential decision about a country, a supplier, or a person.

WHAT IS PUBLISHED, AND HOW IT CAN BE CHECKED

The deliverable is an interactive website that runs in an ordinary browser and can be served from any static host. It carries a world map, a sortable ranking of all 184 scored countries with the unscored cases shown openly, individual country profiles with a phase and domain breakdown, the full indicator and source list, a limitations page, and a simulation page.

The simulation page lets a reader change the balance between the two phases, switch the combining rule from a geometric mean to a plain average, and move a country's domain scores, watching the map and rankings recompute. Its own description is careful: this shows how sensitive the index is to the choices made in building it, and is not a forecast of anything, nor a prediction of what any intervention would achieve.

Every figure shown on the site is read from the published build output rather than entered by hand, and the whole build can be regenerated with a single command from the inputs stored in the repository. The rebuild verifies its own output against the published baseline and stops if anything has drifted unexpectedly. Roughly two-thirds of the data sources can be re-pulled automatically from public interfaces; the remainder require a human to obtain a registration-gated or licence-gated file first, and each of these is listed with its provider, its address, and its licence terms.

The per-indicator source, vintage, coverage, licence, and required citation are recorded for every signal. The codebook that documents which indicators sit in which domain is generated from the code itself, so it cannot drift away from the pipeline it describes. These are small pieces of discipline, and they are the reason the rest of this report could be written from the repository alone.

INTENDED AUDIENCE AND USE

The index is addressed to those working on prevention rather than prosecution: labour ministries and inspectorates, humanitarian and development programmers, procurement and due-diligence teams carrying responsibilities under the UN Guiding Principles, and researchers.

Its stated value is as a triage instrument. It points to where vulnerability and unchecked exploitation hold together, consistently across countries that are otherwise difficult to compare, so that attention can be directed before harm becomes visible. It is a starting point for inquiry, not a verdict on any country.

The project is explicit about the boundary of the intervention material it publishes. Identifying where the structure is sensitive is not the same as predicting what an intervention would achieve. The step from noticing a lever to knowing that pulling it works in a particular place belongs to evaluation on the ground.

Three disciplines govern responsible use: read the structure rather than a body count; read inside the country rather than only the national figure, where the sub-national layer allows it; and read tiers rather than ranks.

14

CONCLUSION

FLSRI addresses a real and well-recognised problem: the places where forced labour is most likely are frequently the places where it is least well recorded, which leaves prevention work waiting for evidence that arrives late or not at all. By measuring conditions rather than cases, the index offers a way to prioritise attention without pretending to count what cannot yet be counted.

Its most substantial contribution is conceptual and procedural rather than numerical. The insistence that a country scores high only where both an exposed population and an unchecked environment are present is a genuine structural claim, encoded in the arithmetic rather than asserted in prose. The refusal to convert missing data into zero, the exclusion of inputs that would predict forced labour from a measure of forced labour, and the decision to publish uncertainty bands wide enough to embarrass the ranking are all choices that make the index harder to over-read.

The honest summary of its current state is that its structure is defensible and its individual country figures are not yet findings. Its Exploitation half rests on one domain that does not measure what it was designed to measure. It under-reads the sponsorship systems that drive some of the most thoroughly documented forced labour in the world, and it says so in the same breath as it publishes the ranks that reflect that gap. It shares a great

deal of its variation with measures of institutional quality, and it demonstrates the bound rather than denying the overlap.

What makes the work useful is that these limits are not concessions extracted from it but statements it makes about itself, in its documentation, in its code comments, and on the face of the published site. An index that tells its reader where not to trust it is more usable than one that does not, and that is the standard by which this prototype should be judged and, in time, improved.

References

SOURCES

- 01 International Labour Organization. Forced Labour Convention, 1930 (No. 29), Article 2(1); Abolition of Forced Labour Convention, 1957 (No. 105); and the 2014 Protocol to Convention No. 29.
- 02 International Labour Organization. ILO Indicators of Forced Labour, first issued 2012, revised edition 2025.
- 03 United Nations. Guiding Principles on Business and Human Rights, endorsed by the UN Human Rights Council in 2011.
- 04 United Nations. Sustainable Development Goal Target 8.7, with SDG indicators 8.3.1 and 16.9.1 used directly by the index.
- 05 World Bank. World Development Indicators, including poverty headcount, agricultural employment, GDP growth volatility, the Gini index, birth registration completeness, and firm bribery incidence.
- 06 International Labour Organization. ILOSTAT, including informal employment share, labour inspector density, collective bargaining coverage, trade union density, and sex-by-sector employment.
- 07 Varieties of Democracy project. V-Dem dataset, freedom of movement and clientelism indicators.
- 08 UNHCR. Refugee population statistics, covering refugees originating from a country, internally displaced persons, and statelessness.
- 09 Uppsala Conflict Data Program. Georeferenced Event Dataset, conflict deaths 2019 to 2023.
- 10 CRED, UCLouvain. EM-DAT, the international disaster database, disaster-affected intensity 2020 to 2024.
- 11 World Bank. Global Findex Database, financial account access and borrowing behaviour.
- 12 Henley and Partners. Henley Passport Index, visa-free destination access.
- 13 ETH Zurich. Ethnic Power Relations dataset, politically excluded ethnic groups.

- 14 United Nations Development Programme. Gender Inequality Index, Human Development Report.
- 15 World Bank. Women, Business and the Law, mobility component.
- 16 Notre Dame Global Adaptation Initiative. ND-GAIN Country Index, vulnerability axis.
- 17 UNCTAD. UNCTADSTAT, merchandise export concentration.
- 18 TRACE International. Bribery Risk Matrix.
- 19 UNODC. Global Report on Trafficking in Persons, forced-labour share of detected victims, 142 countries.
- 20 DEMSCORE. Legal-status data used for the tied-status coding covering 29 countries.
- 21 Basel Institute on Governance. Basel AML Index, FATF effectiveness scores and list standing, used in the Monetization lens only.
- 22 Tax Justice Network. Financial Secrecy Index, used in the Monetization lens only.
- 23 World Bank. Informal Economy Database, shadow economy as a share of official GDP.
- 24 IPUMS-International. Harmonised census microdata behind the sub-national layer, with the Geocoded Disasters dataset for sub-national disaster exposure.
- 25 Walk Free. Global Slavery Index, used as the external prevalence benchmark and for vulnerability dimension alignment.